Homework 1 on Distributed Database Design

Problem 5.2. Consider relation ASG in Figure 5.3. Suppose there are two applications that access ASG. The first is issued at five sites and attempts to find the duration of assignment of employees given their RESPs. Assume that managers, consultants, engineers, programmers, and analysts are located at five different sites. The second application is issued at two sites where the employees with an assignment duration of less than 20 months are managed at one site, whereas those with longer duration are managed at a second site. Derive the primary horizontal fragmentation of ASG using the foregoing information.

EMP

ENO	ENAME	TITLE
E1	J. Doe	Elect. Eng
E2	M. Smith	Syst. Anal.
E3	A. Lee	Mech. Eng.
E4	J. Miller	Programmer
E5	B. Casey	Syst. Anal.
E6	L. Chu	Elect. Eng.
E7	R. Davis	Mech. Eng.
E8	J. Jones	Syst. Anal.

ASG

ENO	PNO	RESP	DUR
E1	P1	Manager	12
E2	P1	Analyst	24
E2	P2	Analyst	6
E3	P3	Consultant	10
E3	P4	Engineer	48
E4	P2	Programmer	18
E5	P2	Manager	24
E6	P4	Manager	48
E7	P3	Engineer	36
E8	P3	Manager	40

PROJ

PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Montreal
P2	Database Develop.	135000	New York
P3	CAD/CAM	250000	New York
P4	Maintenance	310000	Paris

PAY

TITLE	SAL
Elect. Eng.	40000
Syst. Anal.	34000
Mech. Eng.	27000
Programmer	24000

Figure 5.3. Modified Example Database

Problem 5.8. Assume the following view definition

```
CREATE VIEW   EMPVIEW(ENO, ENAME, TITLE, PNO, RESP) =
AS      SELECT   EMP.ENO, EMP. ENAME, EMP. TITLE, ASG.PNO, ASG.RESP
        FROM     EMP, ASG
        WHERE    EMP.ENO=ASG.ENO AND ASG.DUR=24 AND
                 EMP.TITLE="Programmer"
```

is accessed by application q1, located at sites 1 and 2, with frequencies 10 and 20, respectively.

Let us further assume that there is another query q2 defined as

```
SELECT    ENO, DUR
FROM      ASG
```

which is run at sites 2 and 3 with frequencies 20 and 10, respectively. Based on the above information, construct the *use*(*qi, Aj*) matrix for the attributes of both relations EMP and ASG. Also construct the affinity matrix containing all attributes of EMP and ASG. Finally, transform the affinity matrix so that it could be used to split EMP and ASG into two vertical fragments, respectively, using heuristics.

Problem 5.17. Assume the environment of Exercise 5.8. Also assume that 60% of the accesses of query q1 are updates to PNO and RESP of view EMPVIEW and that ASG.DUR is not updated through EMPVIEW. In addition, assume that the data transfer rate between site 1 and site 2 is half of that between site 2 and site 3. Based on the above information, find a reasonable fragmentation of ASG and EMP and an optimal replication and placement for the fragments, assuming that storage costs do not matter here, but copies are kept consistent.

Hint: Consider horizontal fragmentation for ASG based on DUR=24 predicate and the horizontal fragmentation for EMP based on TITLE="Programmer". Also look at the affinity matrix obtained in Example 5.8 for EMP and ASG together, and consider whether it would make sense to perform a vertical fragmentation for ASG.